

# Heterogeneous Cell Population Dynamics: Equation-Free Uncertainty Quantification Computations

Katherine A. Bold, Yu Zou, and Ioannis G. Kevrekidis\*

Department of Chemical Engineering and  
Program in Applied and Computational Mathematics

Princeton University

Princeton, NJ 08544

USA

Michael A. Henson

Department of Chemical Engineering

University of Massachusetts

Amherst, MA 01003

USA

*Submitted to Biophysical Journal*

*February 9, 2008*

## ABSTRACT

We propose a computational approach to modeling the collective dynamics of populations of coupled heterogeneous biological oscillators. In contrast to Monte Carlo simulation, this approach utilizes generalized Polynomial Chaos (gPC) to represent random properties of the population, thus reducing the dynamics of ensembles of oscillators to dynamics of their (typically significantly fewer) representative gPC coefficients. Equation-Free (EF) methods are employed to efficiently evolve these gPC coefficients in time and compute their coarse-grained stationary state and/or limit cycle solutions, circumventing the derivation of explicit, closed-form evolution equations. Ensemble realizations of the oscillators and their statistics can be readily reconstructed from these gPC coefficients. We apply this methodology to the synchronization of yeast

---

\*To whom correspondence should be addressed: [yannis@Princeton.edu](mailto:yannis@Princeton.edu);  
+1-609-258-2818

glycolytic oscillators coupled by the membrane exchange of an intracellular metabolite. The heterogeneity consists of a single random parameter, which accounts for glucose influx into a cell, with a Gaussian distribution over the population. Coarse projective integration is used to accelerate the evolution of the population statistics in time. Coarse fixed-point algorithms in conjunction with a Poincaré return map are used to compute oscillatory solutions for the cell population and to quantify their stability.

## INTRODUCTION

Autonomous oscillations are observed in biological systems ranging in complexity from microorganisms to human beings (1). Often these oscillations are generated at the cellular level through positive feedback loops embedded in gene regulatory or metabolic networks (2). Robust strategies have evolved, based on intracellular coupling of *multiple* oscillators (rather than on a single cellular oscillator subject to degradation and failure). A fundamental feature of such multicellular systems is the synchronization of individual oscillators to produce a coherent overall rhythm (3). Synchronized oscillators are responsible for rhythm generation at second (heartbeat generation), daily (circadian timekeeping) and monthly (menstrual cycles) time scales. Malfunctioning of these interconnected oscillators can produce disastrous consequences such as sudden heart attacks and epileptic seizures.

The development of mathematical models to study the synchronization of coupled oscillators has a long and beautiful history (1). Seminal contributions to the fundamental understanding of synchronization have been made by rigorous mathematical analysis of simple model systems (4, 5). The study of mechanistic models of multicellular biological systems typically involves computational approaches such as dynamic simulation and numerical bifurcation analysis. Difficulties in applying such scientific computing tools are largely determined by population model complexity, which in turn is determined by the complexity of the individual cell model and the number of cells included in the population. Certain applications require both a detailed single cell model and a large ensemble of single cells for meaningful computational study. For instance, a single mammalian circadian oscillator contains multiple interconnected feedforward and feedback loops that have been modeled with up to 73 coupled differential equations (6). Meanwhile, the mammalian circadian system is comprised of approximately 10,000 individual oscillators that communicate via neurotransmitter mediated coupling (7). The development of stochastic simulations to generate meaningful population statistics is possible only if the model ensemble size is sufficiently large. Therefore, there is considerable motivation to develop efficient simulation and bifurcation analysis techniques for large, heterogeneous ensembles of coupled, complex biological oscillators. Previous studies of heterogeneity (8, 9) used idealized phase models (a single “phase” equation for each oscillator); in this work, we show how the approach can be applied to ensembles of realistic limit cycle oscillator models.

Among traditional techniques employed in the study of ensemble statistics of stochastic systems, the most popular one is Monte Carlo simulation (10) (or the cell-ensemble method (11) in the biological context). This approach, however, becomes extremely time-consuming when the

number of simulation realizations is large, as in the case of multicellular coupled oscillator populations. As an alternative, the stochastic Galerkin method for uncertainty quantification (UQ) has been widely used in recent years for solving stochastic ODEs or PDEs. Pioneering work along these lines (12) studied stochastic systems with Gaussian random variables: by viewing randomness as an additional dimension, beyond the space and time dimensions, the dependence of system responses on random parameters is represented in terms of orthogonal polynomial expansions of random variables. In this context (see Appendix A) all orthogonal polynomials of a given order are called Polynomial Chaoses or Homogeneous Chaoses of that order; projections of the system responses onto the Polynomial Chaos (PC) coefficients evolve deterministically, and equations for their evolution can in principle be obtained, and then solved, by applying a Galerkin projection. The method has been applied for uncertainty quantification purposes in various physical and engineering systems including structures with random properties (12), porous media (13), fluid dynamics (14) and chemical reactions (15). In (16), the method was extended: *generalized* Polynomial Chaos (gPC), applicable to a variety of continuous and discrete probability measures, was proposed based on the Askey scheme. In addition to representations of orthogonal polynomials, this method was also improved in other directions, in the form of piecewise (h-refinement) representations (17) and wavelet expansions (18).

One advantage of the stochastic Galerkin method, as compared to direct Monte Carlo simulation, is that it can reduce a stochastic system to a deterministic one with (often significantly) fewer degrees of freedom, thus accelerating computation and saving data storage space. In order to apply the stochastic Galerkin method, however, one must derive equations for the temporal evolution of gPC coefficients either explicitly or through a pseudospectral approach (e.g. (19)) and develop a new code for the solution of these equations. To circumvent this additional effort, Equation-Free (EF) methods (20, 21, 22) have been utilized recently to quantify propagation of uncertainty in a stochastic system by evolving gPC coefficients of random solutions using the system dynamic simulator in a *nonintrusive way*, that is, without deriving the corresponding explicit gPC evolution equations (23). Within this multiscale equation-free framework, the original stochastic dynamics code is viewed as a fine-level simulator, while the (unavailable) ODEs for the time-evolution of the gPC coefficients are viewed as a coarse-grained system model. These equation-free algorithms are built based on protocols that enable communication between different levels of system description; the *lifting* protocol translates coarse-grained initial conditions to one or more consistent fine scale initial conditions; the *restriction* protocol computes the coarse-grained description (values of the coarse variables, “observables”) of fine scale system configurations. The success of this class of methods relies on the assumption that closed evolution equations for the dominant (low-order) gPC coefficients exist *in principle*, even if they are not explicitly available.

The purpose of this paper is to demonstrate that efficient simulation strategies for large populations of coupled biological oscillators can be developed by utilizing these equation-free uncertainty quantification (EF-UQ) based methods. A six-dimensional cellular model of yeast glycolytic oscillations (24, 25) is used to study synchronization of 1,000 *heterogeneous* cells in a well mixed environment. The random variable (which characterizes the cell population heterogeneity) is chosen as the glucose influx for each cell. In the context of EF-UQ, we demonstrate coarse projective integration, which accelerates temporal simulation of the cell population dy-

namics, and coarse fixed-point computation combined with Poincaré return maps to efficiently converge on limit cycle solutions, corresponding to synchronous population oscillations. Limits of the applicability of the procedure are discussed, including an extension of the basic methodology to handle loss of synchronization when isolated outlier cells “detach” from the main coherent population and develop individual oscillatory characteristics.

## A MECHANISTIC SYNCHRONIZATION MODEL FOR YEAST GLYCOLYTIC OSCILLATIONS

The yeast *Saccharomyces cerevisiae* exhibits autonomous oscillations with a period of approximately one minute when grown under anaerobic conditions (26, 27, 28). Similar oscillations have been observed in other yeast strains (29, 30) as well as algae (31), muscle (32), heart (33) and tumor (34) cells. Yeast studies suggest that an autocatalytic reaction involving the glycolytic enzyme phosphofructokinase is the main cause of oscillations at the single cell level. Therefore, the observed limit cycle behavior has been termed glycolytic oscillations. Additional experimental work has focused on characterizing the intercellular mechanisms involved in the synchronization of individual yeast cell oscillations (35). Typically, oscillations at the cell population level are observed by continuous monitoring of the average intracellular NADH concentration using fluorometry. Experiments with *Saccharomyces cerevisiae* grown in anaerobic stirred cuvettes suggest that secreted acetaldehyde is the key signaling molecule in the synchronization mechanism (36, 37).

A number of simple cell models have been developed to capture the glycolytic oscillation mechanism in yeast (38, 39, 40, 41). Cell models based on more detailed descriptions of the glycolytic reaction pathway also have been proposed (42, 43). Small ensembles of single cell models have been used to investigate the synchronization phenomenon (38, 41, 43). In this paper, we use a single cell model of intermediate complexity (25) to demonstrate our computational framework for simulating large populations of coupled biological oscillators. Our cell ensemble model is based on an intracellular coupling mechanism involving the transport of acetaldehyde across the cell membrane (24, 25).

A single cell in the population is described by the following differential equations:

$$\frac{dS_{1,i}}{dt} = J_{0,i} - v_{1,i} = J_{0,i} - k_1 S_{1,i} A_{3,i} \left[ 1 + \left( \frac{A_{3,i}}{K_I} \right)^q \right]^{-1} \quad (1)$$

$$\frac{dS_{2,i}}{dt} = 2v_{1,i} - v_{2,i} - v_{6,i} = 2k_1 S_{1,i} A_{3,i} \left[ 1 + \left( \frac{A_{3,i}}{K_I} \right)^q \right]^{-1} - k_2 S_{2,i} (N - N_{2,i}) - k_6 S_{2,i} N_{2,i} \quad (2)$$

$$\frac{dS_{3,i}}{dt} = v_{2,i} - v_{3,i} = k_2 S_{2,i} (N - N_{2,i}) - k_3 S_{3,i} (A - A_{3,i}) \quad (3)$$

$$\frac{dS_{4,i}}{dt} = v_{3,i} - v_{4,i} - J_i = k_3 S_{3,i} (A - A_{3,i}) - k_4 S_{4,i} N_{2,i} - J_i \quad (4)$$

$$\frac{dN_{2,i}}{dt} = v_{2,i} - v_{4,i} - v_{6,i} = k_2 S_{2,i} (N - N_{2,i}) - k_4 S_{4,i} N_{2,i} - k_6 S_{2,i} N_{2,i} \quad (5)$$

$$\frac{dA_{3,i}}{dt} = -v_{1,i} + 2v_{3,i} - v_{5,i} = -2k_1 S_{1,i} A_{3,i} \left[ 1 + \left( \frac{A_{3,i}}{K_I} \right)^q \right]^{-1} + 2k_3 S_{3,i} (A - A_{3,i}) - k_5 A_{3,i} \quad (6)$$

where the index  $i$  denotes the cell. The pathway model accounts for glucose flux into the cell ( $J_0$ ), metabolism of glucose to produce intracellular glycerol, ethanol and a combined acetaldehyde/pyruvate pool (hereafter called acetaldehyde), acetaldehyde flux out of the cell ( $J$ ), and degradation of extracellular acetaldehyde by cyanide. Glycolytic intermediates modeled are intracellular glucose ( $S_1$ ), the glyceraldehyde-3-phosphate/ dihydroxyacetonephosphate pool ( $S_2$ ), 1,3-bisphosphoglycerate ( $S_3$ ) and intracellular acetaldehyde ( $S_4$ ). Each co-metabolite pair is assumed to be conserved with  $A$  and  $N$  denoting the constant concentrations of the ADP/ATP and  $\text{NAD}^+/\text{NADH}$  pools, respectively. Therefore, only the NADH ( $N_2$ ) and ATP ( $A_3$ ) concentrations are treated as independent variables. The intracellular reaction rates  $v_2$ – $v_6$  depend linearly on the metabolite and co-metabolite involved in each reaction, while the acetaldehyde degradation rate, denoted  $v_7$ , depends linearly on the extracellular acetaldehyde concentration. Individual cell oscillations are attributable to the nonlinear term in the reaction rate  $v_1$  that accounts for ATP inhibition.

The net flux of acetaldehyde from the  $i$ -th cell into the extracellular environment is modeled as  $J_i = \kappa(S_{4,i} - S_{4,ex})$  where  $S_{4,ex}$  is the extracellular acetaldehyde concentration and  $\kappa$  is a coupling parameter related to the cell permeability. A mass balance on extracellular acetaldehyde is derived under the assumption that the volume fraction of cells relative to the total medium volume ( $\varphi$ ) remains constant as the total number of cells  $M$  is varied:

$$\frac{dS_{4,ex}}{dt} = \frac{\varphi}{M} \sum_{i=1}^M J_i - v_7 = \frac{\varphi}{M} \sum_{i=1}^M \kappa(S_{4,i} - S_{4,ex}) - kS_{4,ex} \quad (7)$$

where  $k$  is the kinetic constant of the acetaldehyde degradation reaction. The chosen parameter values (24) produce an asymptotic solution in which all cells are synchronized regardless of the cell number. The total number of differential equations ( $n$ ) in the cell ensemble model increases linearly with the number of intracellular metabolites (6) and the number of cells ( $M$ ):  $n = 6M + 1$ . Unless otherwise stated, the following simulations involve 1000 cells:  $n = 6001$ .

## APPLICATION OF EF-UQ COMPUTATION TO YEAST GLYCOLYTIC OSCILLATIONS

### Polynomial chaos representation of cell properties

Our cell ensemble model of yeast population dynamics consists of a large set of coupled nonlinear ODEs. The intracellular metabolite concentrations in such problems are in general random variables; many different sources of randomness exist, from intrinsic kinetic fluctuations, to population heterogeneity, to randomness in the initial conditions. In our particular case we consider the randomness arising from *population heterogeneity*; there is a *single* random parameter, the glucose flux  $J_0$ :  $J_0 = \bar{J}_0 + \sigma_J \xi(\omega)$ , where  $\xi$  is normal over the sampling space  $\Omega$ . This simple choice of a normally distributed uncertainty is made for illustration/validation purposes; the procedure is directly applicable to different distributions of uncertainty, as we will discuss below. As

our heterogeneous population evolves, each intracellular concentration evolves, and so therefore do the corresponding concentration distributions over the population. The “obvious” collective variables for evolving distributions are the first few moments of the distribution: mean, variance etc.; it might, at first sight, appear that good coarse-grained observables of our population state would be these first few moments of the individual intracellular concentration distributions. These moments, however, do not take into account *correlations* across the distributions (8); in our heterogeneous population it is not enough to know *how many* cells have a certain intracellular metabolite concentration – we must also know *which* cells have this concentration: the cells with higher or with lower intrinsic values of  $J_0$  ?

We have observed in our simulations (and the same phenomenon has been documented in different heterogeneous coupled oscillator contexts (9)) that, after a typical initialization, two distinct phases are observed in the dynamics. During an initial, *relatively fast* phase, strong correlations develop between the distributions of the various intracellular concentration values; this is followed by a second, *long term* phase, during which the distributions evolve, but with the correlations “locked in”. In effect, these correlations appear to be strongly related to the heterogeneity of the population - cells with different  $J_0$  exhibit systematically different concentration patterns. Figure 1 shows the dependence of concentrations of NADH and ATP on the population heterogeneity parameter (the random variable  $J_0$ ) at three instantaneous states. There is clearly a relation/dependence between the intracellular metabolite concentrations and the “identity” of the cell (the value of  $J_0$ ).

If the simulation is continued from the exact state where it was interrupted, these correlations remain in place (see the red curves in Figures 2 and 3). If we now create new, *artificial* initial conditions, where leading moments of the distributions of individual intracellular concentrations are retained *but in which the correlations are deliberately scrambled* then the system will quickly move away from the synchronized oscillation in a violent transient (see the cyan curves in Figures 2 and 3), and will take a long time to return to it, rebuilding the correlations in the process. These numerical experiments suggest that strong correlations between cell heterogeneity and cell behavior get established during initial stages of the population response, and are then retained in the long term dynamics. Based on this observation, and on the functional dependence of intracellular concentrations on our random variable  $\xi$  clearly apparent in Figure 1, we will assume that all intracellular concentrations across the population can, in the long-term dynamics, be expressed as (unknown) functions of the same random variable.

Denoting

$$\mathbf{x}(\xi(\omega), t) = (S_1(\xi(\omega), t), S_2(\xi(\omega), t), S_3(\xi(\omega), t), S_4(\xi(\omega), t), N_2(\xi(\omega), t), A_3(\xi(\omega), t))^T,$$

we will represent  $\mathbf{x}$  in terms of a truncated Polynomial Chaos expansion of  $\xi$

$$\mathbf{x}(\xi, t) = \sum_{j=0}^P \mathbf{x}_c^j(t) \Psi_j(\xi). \quad (8)$$

The fine scale state is the 6001-long vector of dependent variables (the 6000 intracellular concentrations plus one extracellular one) in our detailed set of coupled ODEs. Our coarse-grained observables are the gPC truncation coefficients,  $\mathbf{x}_c^j(t) = (x_{1,c}^j, x_{2,c}^j, x_{3,c}^j, x_{4,c}^j, x_{5,c}^j, x_{6,c}^j)^T$ . The



*lifting* step, the construction of ensemble realizations of intracellular concentrations (fine-level states) consistent with a particular set of values of gPC coefficients (coarse-level observables) is performed through

$$\mathbf{x}(\xi_i, t) = \sum_{j=0}^P \mathbf{x}_c^j(t) \Psi_j(\xi_i), \quad i = 1, 2, \dots, M. \quad (9)$$

Here  $M(= 1000)$  is the total number of cells, the truncation level  $P$  is set to 3, and  $\{\Psi_j\}$  are orthonormal Hermite polynomials (44) for the case of normally distributed  $\xi$ . We reiterate that different types of polynomials can be used for random variables obeying different distributions. The total number of our *coarse* model states is thus  $4 \times 6 + 1 = 25$  (the deterministic extracellular concentration  $S_{4,ex}$  is counted here as a single coarse observable); this is clearly much less than the number of internal states,  $6M + 1$ , in the original cell dynamics.

Obtaining the coarse grained observables from a detailed state constitutes the *restriction* step; our restriction protocol consists of the inner product, Eq. 12, which can be performed in one of two ways: (i) we can discretize the integral to approximate the inner product, i.e.,  $\langle \mathbf{x}, \Psi_i \rangle = 1/M \sum_{j=1}^M \mathbf{x}(\xi_j) \Psi_i(\xi_j)$ ; (ii) we can perform a simple least squares fitting to find  $\mathbf{x}_c^j$  such that an  $L^2$  norm  $\|\mathbf{x}(\xi, t) - \sum_{j=0}^P \mathbf{x}_c^j(t) \Psi_j(\xi)\|$  is minimized. We used the second implementation to obtain gPC coefficients in this work.

In what follows, we will demonstrate two distinct ways of using the EF-UQ methodology in order to accelerate population computations for our model problem. We will first demonstrate direct simulation acceleration through coarse projective integration. We will then demonstrate the accelerated computation and coarse-grained stability analysis of synchronized population oscillations through matrix-free fixed point computation and eigenvalue approximation; these synchronized limit cycles will be computed as fixed points of a coarse Poincaré map. By doing so, we demonstrate that our multiscale toolkit provides a generally applicable and computationally efficient framework for dynamic simulation and analysis of heterogeneous populations of cellular oscillators.

## Full direct simulation and coarse projective integration

The full direct simulation consists of integrating of  $6M + 1$  differential equations ( $M$  is the number of cells, 6 internal states for each cell, and 1 extracellular variable). The package ODETools is used in Matlab for this simulation, with variable step-size chosen for relative error tolerance of  $1 \times 10^{-9}$  and absolute error tolerance of  $1 \times 10^{-12}$ . Figure 4 shows the time history of such a full simulation of an ensemble of cells over one complete period for the distribution of parameter  $J_0$  having mean 2.3 and standard deviation  $1 \times 10^{-3}$ . At discrete moments in time, the values of the full ensemble are projected to the gPC basis (the full state is *restricted* onto coarse observables), also shown in Figure 4. At the fine scale, the concentration of each intracellular metabolite oscillates and, at the end of an oscillation, returns exactly to its starting value. At the coarse, macroscopic level, it is the gPC coefficients that return to themselves. This is, in effect, saying the distributions of the species concentrations return to their initial values.

We now explain the projective integration procedure and compare its results to the direct simulation. Starting at  $t = t_0$  from a given coarse initial condition (values of the first few gPC

coefficients, the observables), we generate fine-level model states by the *lifting* procedure Eq. 9. We use the original, full dynamic simulation code to evolve the fine-level description for an initial time interval  $t_{heal}$ ; we continue the full direct simulation over three successive time intervals of time  $\delta t$ . At the end of each of these intervals, the coarse observables are obtained by *restriction*. The temporal derivatives of the coarse observables can then be estimated (here for convenience we use least squares, but maximum-likelihood based methods are more appropriate (45)); these local time derivatives are then used to *project* the values of the coarse variables at a time  $5\delta t$  further in the future. Using these predicted values we begin the same cycle of brief healing, detailed simulation observed by restriction, estimation of local time derivatives and projection of the coarse behavior into the future. Depending on the coarse projection algorithm used in the forward-in-time projection, more (or fewer) restrictions from the short burst of full simulation may be needed. The particular values of  $t_{heal}$  and  $\delta t$  can also vary from those used here ( $t_{heal} = 5 \times 10^{-3}$ ,  $\delta t = 5 \times 10^{-3}$ ), and their on-line optimal selection is problem-dependent.

Figure 5 demonstrates the acceleration of the full simulation through a coarse projective forward Euler algorithm. Note that for every  $4\delta t$  of full direct simulation, we project  $5\delta t$  into the future – only 44.44% of the work is necessary this way. This type of computational savings enables the simulation of larger cell ensembles, allowing more accurate reconstructions of population statistics. As discussed in (46) the method can provide significantly larger savings if a separation of time scales exists in the problem; for linear problems this can be readily seen as a gap between a few leading, slow, and the remaining many, fast eigenvalues of the system. Different coarse projective algorithms (e.g. projective Runge-Kutta, or even *implicit* projective algorithms) can also be used; linking them to modern estimation techniques and extending them to account for adaptive projective step size selection is the subject of ongoing research (see (47) as well as (48)).

## Coarse-grained limit cycle computations

**Limit cycle computation.** Beyond direct simulation, which asymptotically approaches *stable* limit cycles, periodic orbits are located by solving boundary value problems in time. In particular, they can be located as fixed points of a *Poincaré map* (see, for example, the textbook (49)). A *Poincaré Section*,  $S$ , is a hypersurface (often a hyperplane) crossing a limit cycle transversely at an (isolated) point. The Poincaré map,  $P : S \rightarrow S$ , is a return map defined by

$$P(\mathbf{x}) = \phi_t(\mathbf{x}) \in S,$$

for  $\mathbf{x} \in S$ ,  $t$  the smallest positive time for which  $\phi_t(\mathbf{x}) \in S$ , and  $\phi_t$  the time  $t$  flow map. Note that it is not necessary to know *a priori* the period of the limit cycle when defining the Poincaré map. Let  $P_f$  be the Poincaré map for the full simulation and  $P_c$  be the Poincaré map for the coarse simulation. The two maps are related by

$$P_c = R \circ P_f \circ L$$

for  $R, L$  the restricting and lifting operators and  $\circ$  is composition. We use the Newton-Krylov GMRES method to find solutions of  $P_c(\mathbf{X}_c) = \mathbf{X}_c$  (50). This implementation of Newton iteration



does not explicitly require the computation of the Jacobian of the equation to be solved (often computed in shooting methods through integration of the variational equations). The *action* of this Jacobian along sequentially selected directions (directional derivatives) is estimated from the results of simulations starting at appropriately chosen nearby initial conditions. Since this only requires simulations of the problem, and the linear equations are solved without ever assembling the relevant Jacobian, this is a *matrix-free* implementation.

Note that the Poincaré return map can be constructed, not only by direct simulation, but also by projective integration. That is, projective integration can be implemented to accelerate the computations of the Poincaré map itself. The limit cycle in Figure 5 was found by solving this fixed point problem.

**Limit cycle stability.** For multiscale problems with limit cycles, we can discuss stability at the fine-scale and also at the coarse-grained level. We describe here limit cycle stability computations at both levels, and it is interesting that the results are essentially the same – the coarse-grained stability computations reveal the same information as the (computationally intensive) fine-scale computations.

Let  $T$  be the period of the limit cycle, and let  $\mathbf{x}_0$  be a point on the limit cycle. The time  $T$  flow map is  $\phi_T(\mathbf{x}_0) = \mathbf{x}(t; \mathbf{x}_0)$ . Eigenvalues (*Floquet multipliers*) of the matrix  $D\phi_T(\mathbf{x}_0)$  (the Jacobian of  $\phi_T$  at  $\mathbf{x}_0$ , the so-called *monodromy matrix*) quantify the linearized stability of the limit cycle. The time  $T$  coarse flow map is  $\Phi_T = \Phi_T(\mathbf{X}_c)$ , see Eq. 19. The matrix  $D\Phi_T(\mathbf{X}_c)$  can be approximated by finite differences,

$$\left[ \frac{\Phi_T(\mathbf{X}_c + \varepsilon \mathbf{e}_1) - \mathbf{X}_c}{\varepsilon}, \frac{\Phi_T(\mathbf{X}_c + \varepsilon \mathbf{e}_2) - \mathbf{X}_c}{\varepsilon}, \dots \right]$$

for  $\varepsilon \ll 1$ . The eigenvalues of  $D\Phi_T(\mathbf{X}_c)$  are shown in Figure 6, and the leading ones are listed in Table 1. Note that the number 1 is an eigenvalue (corresponding to time translational invariance *along* the limit cycle at the point  $y_0$ ). This invariance gives rise to a neutrally stable direction; all limit cycles possess this neutral eigenvalue at 1. The magnitudes of the remaining leading eigenvalues of the limit cycle are less than 1 (they lie inside the unit circle). Therefore, the limit cycle is stable.

Eigenvalues of the fine-scale monodromy matrix reveal the stability of the full limit cycle. Divided differences could in principle be used to approximate the full  $6001 \times 6001$  Jacobian of the fine scale problem; instead, to obtain this matrix we integrated the variational equations (see, for example, the textbook (49)). Let  $\dot{\mathbf{x}} = f(\mathbf{x})$ ; then the variational equations along the solution  $\mathbf{x}(t, \mathbf{x}_0)$  are

$$U'(t, \mathbf{x}_0) = Df(\mathbf{x}(t, \mathbf{x}_0)) \cdot U(t, \mathbf{x}_0),$$

where  $U(t)$  is a  $(6M+1) \times (6M+1)$  matrix and the initial condition  $U(0)$  is the identity matrix. This gives

$$D\phi_T^f(\mathbf{x}_0) = U(T, \mathbf{x}_0).$$

The “fine” eigenvalues are shown in Figure 6, and the leading ones are in Table 1.

It can be shown that eigenvalues of the coarse and fine-scale monodromy matrices are theoretically the same (if sufficiently many gPC coefficients are kept), and the coincidence of the

eigenvalues is seen in the table. Analysis of the fine-scale model considered here is computationally tractable; for larger problems, however, a full analysis may not be possible, and the coarse stability analysis suffices to characterize stability of the fine-scale limit cycles. For instance, these methods could be profitably applied to large models of coupled biological oscillators involved in yeast respiratory oscillations (51) and mammalian circadian rhythm generation (52).

## When coarse-graining fails: “rogue” oscillators

When the variance of  $J_0$  is small, NADH concentrations ( $N_2$ ) in the cells across the population are narrowly distributed; an overall synchronized solution for the entire population prevails, and our coarse-graining methods are successful. When the variance of  $J_0$  increases, however, the amplitudes of the  $N_2$  oscillations across the cell population will separate significantly. For certain combinations of mean and variance of  $J_0$ , one or more cells will appear to oscillate “freely” in amplitude. An illustration of this, using a 50 cell ensemble as the full direct simulation, and setting the mean and standard deviation of  $J_0$  to 2.1 and 0.08, respectively is seen in Figure 7. The cause of this single “free, oscillating” cell can be clearly rationalized based on the relatively large variance of  $J_0$ : the values of  $J_0$  across the population spread *beyond* the parameter point at which the single cell dynamics undergoes a Hopf bifurcation. The strong correlation between individual cell  $J_0$  values and their detailed states (intracellular concentrations) is retained for *most* of the cells, but lost for the outlying “rogue” oscillatory cell. As  $\sigma_J$  continues increasing, the strong correlations that allowed us to use a gPC expansion for the collective behavior fail even more dramatically - see Figure 7, which is quite representative of dynamics in populations with large variance of the glucose influx. We expect that such strong discontinuities in the relation between heterogeneity and detailed state (reminiscent, at some level, of Gibbs phenomena) will, in general, be encountered in probabilistic problems involving strong nonlinearity and bifurcations of the single oscillator behavior as the heterogeneity parameter(s) is varied. Clearly, a few gPC coefficients are no longer good observables of the collective system state. The recent literature includes some efforts in resolving such cases using a wavelet-based chaos expansion (18) or piecewise Polynomial Chaos (17); yet the problem remains an open subject for future research.

Here we will limit ourselves to the case of a *single* rogue oscillator, shown above. A simple and rational way to tackle this problem is to employ gPC coefficients to describe the oscillators that are “clumped together” with smaller oscillation amplitudes, and use an additional, *different* set of variables to describe intracellular metabolite concentrations for the single “freely oscillating” cell. More specifically, for a total number  $M$  of cells which include one “free” cell, the coarse observables are comprised of  $\mathbf{x}_c^0, \mathbf{x}_c^1, \dots, \mathbf{x}_c^P$  (obtained by restricting intracellular concentrations of  $M - 1$  clumped cells through the least-square fitting method mentioned earlier), the extracellular concentration  $S_{4,ex}$  and six intracellular concentrations  $S_1, S_2, S_3, S_4, N_2, A_3$  representative of the free cell. When lifting is now implemented, only intracellular concentrations of  $M - 1$  cells are generated from  $\mathbf{x}_c^0, \mathbf{x}_c^1, \dots, \mathbf{x}_c^P$  through Eq. 9. The intracellular concentrations  $S_1, S_2, S_3, S_4, N_2, A_3$  for the free cell are part of the coarse-grained description. Therefore, the full direct simulation is again characterized by  $6M + 1$  variables, but the number of observables of the reduced, coarse-grained problem increases to  $4 \times 6 + 1 + 6 = 31$  if the first four leading order gPC coefficients are retained ( $P = 3$ ).

The initial condition on coarse observables for the projective integration is found by using the equation-free fixed-point algorithm in conjunction with the Poincaré map on the 31-dimensional space. Coarse projective integration for this new, 31-dimensional coarse grained system is illustrated in Figure 5. Our coarse initial condition is the restriction of a point on the detailed, fine-scale limit cycle solution; starting there, we use short bursts of full direct simulation to estimate the time derivatives of our 31 coarse observables, and then coarse projective forward Euler to project their values forward in time.

The phase portrait of the coarse-grained limit cycle, projected on the leading order gPC coefficient of the concentration of NADH of the  $M - 1$  clumped cells and the concentration of ATP in the free cell, is shown in Figure 5. Reasonable numerical agreement with the full direct oscillation (observed on the same variables) prevails; even in the presence of one (more generally, of a few) free cell(s), choosing a good set of coarse-grained variables allows us to accelerate the computations of the long-term system dynamics. Of course, this approach requires *a priori* identification of the rogue oscillating cells to construct the coarse variables. Clearly more research is needed towards the selection of good coarse observables for such problems.

## SUMMARY and CONCLUSIONS

Equation-Free Uncertainty Quantification methods were used in this paper to accelerate the computer-aided analysis of the dynamics of *heterogeneous* ensembles of coupled biological oscillators; in particular, coarse-grained computations of synchronized population-wide limit cycles and their stability was demonstrated for an ensemble of yeast glycolytic oscillators coupled by membrane exchange of intracellular acetaldehyde. The feasibility of the EF UQ in describing certain particular situations (where one –or a few– oscillator(s) move freely in amplitude, distinguished from the “bulk” of the population) was also demonstrated.

This paper contains only representative “proof of concept” computations. There is a clear necessity for extensive numerical analysis of the schemes illustrated, including adaptive step-size selection, error estimation and control. Different restriction schemes need to be devised when the relation between heterogeneity and behavior becomes nonsmooth or discontinuous in the “randomness direction(s)”. Different *lifting* schemes (53) exploiting a non-explicit separation of time scales may reduce the number of gPC coefficients required for an accurate reduced description. Furthermore, a wealth of data-mining techniques currently under development (and, in particular, the diffusion map approach (54, 55)) holds the promise of extracting low-dimensional parameterizations of high-dimensional data based on graphs constructed on simulation data and the eigenfunctions of diffusion operators on these graphs. It would be interesting to explore the performance of such techniques when simple gPC observables fail, as in the case of the discontinuities and “multiple oscillator clumps” mentioned above.

# APPENDIX A: SOME BASIC UNCERTAINTY QUANTIFICATION (UQ) AND EQUATION FREE UNCERTAINTY QUANTIFICATION (EF-UQ) ISSUES

## On polynomial chaos expansion of random variables and processes

Orthogonal polynomials of random variables with an arbitrary probability measure (Gaussian, uniform, Poisson, binomial, ...) are called generalized Polynomial Chaos (gPC) (16). Any functional vector,  $\mathbf{x}(\omega)$ , of random variables,  $\boldsymbol{\xi}(\omega) (= (\xi_1(\omega), \dots, \xi_n(\omega))^T)$ , defined over a probability space  $(\Omega, \mathcal{F}, P)$ , can be expressed in terms of a gPC expansion,

$$\mathbf{x}(\omega) = \sum_{i=0}^{\infty} \mathbf{x}_c^i \Psi_i(\boldsymbol{\xi}(\omega)), \quad (10)$$

where  $\Psi_i(\boldsymbol{\xi}(\omega))$  is the  $i$ th generalized Polynomial Chaos which admits the following orthogonality properties,

$$\langle \Psi_i, \Psi_j \rangle = \begin{cases} \langle \Psi_i^2 \rangle, & \text{if } i = j, \\ 0, & \text{if } i \neq j, \end{cases} \quad (11)$$

and  $\mathbf{x}_c^i$  is the corresponding coefficient of  $\Psi_i(\boldsymbol{\xi}(\omega))$ , determined by

$$\mathbf{x}_c^i = \frac{\langle \mathbf{x}(\omega), \Psi_i(\boldsymbol{\xi}(\omega)) \rangle}{\langle \Psi_i^2 \rangle}. \quad (12)$$

In the above equations, the inner product  $\langle \cdot, \cdot \rangle$  is defined by

$$\langle f(\boldsymbol{\xi}), g(\boldsymbol{\xi}) \rangle = \int_{\Gamma} f(\boldsymbol{\xi})g(\boldsymbol{\xi})p(\boldsymbol{\xi})d\boldsymbol{\xi}, \quad \boldsymbol{\xi} = (\xi_1(\omega), \dots, \xi_n(\omega))^T, \quad (13)$$

where  $p(\boldsymbol{\xi})$  is the joint probability measure of  $\boldsymbol{\xi}$  and  $\Gamma$  the support of  $p(\boldsymbol{\xi})$ .

In the case that  $\mathbf{x}$  is a random field or process having the form  $\mathbf{x}(\omega, s)$  or  $\mathbf{x}(\omega, t)$  where  $s$  and  $t$  are, respectively, spatial and time coordinates, the projections of  $\mathbf{x}$  onto the Polynomial Chaos,  $\mathbf{x}_c^i$ , must admit a form that depends on these spatial or time coordinates as well. In our illustrative example the gPC coefficients evolve in (and thus also depend on) time.

## On the stochastic Galerkin method

The stochastic Galerkin method aims at quantifying propagation of uncertainty in dynamical systems. In this method, solutions of stochastic systems are first expressed in terms of a finite linear combination of generalized Polynomial Chaos. The error resulting from the finite-term expansion is then required to be orthogonal to test functions, which are normally chosen to be the same as the generalized Polynomial Chaos. A coupled system of equations for the gPC coefficients can thus be derived and solved (12). Probability distributions and statistical moments of the solutions can be computed from the gPC coefficients subsequently.

In particular, let the states of a stochastic system be represented by a vector  $\mathbf{x}(\omega, t) : \Omega \times \mathbb{R} \rightarrow \mathbb{R}^N$ , where  $\omega$  is an element in the sampling space  $\Omega$ . This system state is governed by the differential equation

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x}, \boldsymbol{\xi}(\omega)), \quad \mathbf{x}(\omega, 0) = \mathbf{x}_0(\omega). \quad (14)$$

The solution of the above equation can be approximated by a finite expansion in the form of Eq. 10,

$$\mathbf{x}(\omega, t) = \sum_{i=0}^P \mathbf{x}_c^i(t) \Psi_i(\boldsymbol{\xi}(\omega)). \quad (15)$$

By applying a Galerkin projection, equations governing the gPC coefficients  $\mathbf{x}_c^i(t)$  are obtained as

$$\frac{d\mathbf{x}_c^i}{dt} = \frac{1}{\langle \Psi_i^2(\boldsymbol{\xi}) \rangle} \langle \mathbf{f}(\sum_{i=0}^P \mathbf{x}_c^i(t) \Psi_i(\boldsymbol{\xi})), \Psi_i(\boldsymbol{\xi}) \rangle, \quad i = 0, 1, \dots, P, \quad (16)$$

with  $\mathbf{x}_c^i(0) = \frac{\langle \mathbf{x}_0, \Psi_i(\boldsymbol{\xi}) \rangle}{\langle \Psi_i^2(\boldsymbol{\xi}) \rangle}$ . The above equation can be rewritten as

$$\frac{d\mathbf{X}_c}{dt} = \mathbf{H}(\mathbf{X}_c). \quad (17)$$

Here  $\mathbf{X}_c = (\mathbf{x}_c^0, \mathbf{x}_c^1, \dots, \mathbf{x}_c^P)^T$  and  $\mathbf{H} = (\mathbf{h}_0, \mathbf{h}_1, \dots, \mathbf{h}_P)^T$ , where

$$\mathbf{h}_i(\mathbf{X}_c) = \frac{1}{\langle \Psi_i^2(\boldsymbol{\xi}) \rangle} \langle \mathbf{f}(\sum_{i=0}^P \mathbf{x}_c^i(t) \Psi_i(\boldsymbol{\xi})), \Psi_i(\boldsymbol{\xi}) \rangle, \quad i = 0, 1, \dots, P.$$

If  $\frac{d\mathbf{X}_c}{dt} = \mathbf{0}$  in the long-time limit, then Eq. 17 has a steady state, which can then be used to recover the probability distribution of the random steady state of Eq. 14. The stochastic Galerkin method can provide an effective reduction of a stochastic model if a relatively small truncation in Eq. 15 above is sufficiently accurate. The corresponding deterministic model will then be considerably easier to simulate and analyze than large numbers of realizations in a Monte Carlo simulation of the original dynamics.

## On Equation-Free methods and their application in uncertainty quantification

The basic building block of equation-free methods (20, 21, 22) is the *coarse time-stepper*, which consists essentially of three components: *lifting*, *micro-simulation*, and *restriction*. *Lifting* is a procedure to transform a coarse-level state to its fine-level counterpart and *restriction* the converse of lifting. By employing the model states  $\mathbf{x}(t)$  in Eq. 14 as the fine-level system state vector, and their low-order gPC coefficients  $\mathbf{X}_c$  as the coarse-level observables, Equation-Free methods can be used to numerically study the behavior of gPC coefficients without needing closed form ODEs for their evolution, Eq. 17 (23). The *lifting* and *restriction* protocols in our context are

then Eq. 15 and Eq. 12, respectively. The *micro-simulation* is just the simulation of the original large coupled ODE system Eq. 14. Assuming that the long-term coarse-grained dynamics in gPC space lie on a low-dimensional, attracting *slow manifold*, one can use *coarse projective integration* to accelerate the computation of successive coarse-level system states, i.e., low-order gPC coefficients  $\mathbf{x}_c^i(t_j), i = 0, 1, \dots, P, j = 0, 1, \dots, M$ , through *restriction* of (one or more) short bursts of fine scale simulation as follows: Temporal derivatives  $\frac{d\mathbf{x}_c^i}{dt}$  at  $t_M$  are estimated by least-squares fitting of the short-time gPC coefficient evolution  $\mathbf{x}_c^i(t_j), j = M - k, M - k + 1, \dots, M, k < M$ . These gPC coefficients in conjunction with their locally estimated temporal derivatives are then used to *extrapolate* (in effect, integrate the gPC coefficients numerically in time) over a relatively large coarse time interval  $T$ . For instance, if the coarse forward Euler projective integration is used, then the predicted gPC coefficients at a later time  $t_M + T$  are obtained by

$$\mathbf{x}_c^i(t_M + T) \approx \mathbf{x}_c^i(t_M) + \frac{d\mathbf{x}_c^i}{dt}(t_M)T, \quad i = 0, 1, \dots, P. \quad (18)$$

These *projected in time* gPC coefficients  $\mathbf{x}_c^i(t_M + T)$  are *lifted* again to the full fine state level, and a new short burst of micro-simulation is initiated. The procedure is repeated until a desired time limit is reached. Issues of time-step selection, estimation and error control are important, and often “the devil lies in these details”; discussing these numerical analysis issues, however, is not the aim of this brief exposition, and we refer the reader to (21, 46, 47, 56).

If equation Eq. 17 possesses a steady state  $\mathbf{X}_c^s$ , then  $\mathbf{X}_c^s$  must satisfy an integral form of Eq. 17 given by

$$\mathbf{X}_c^s = \mathbf{X}_c^s + \int_{t_0}^{t_0+T} \mathbf{H}(\mathbf{X}_c)dt \quad (19)$$

where  $\mathbf{X}_c(t_0) = \mathbf{X}_c^s$ . The right-hand side of Eq. 19 can be viewed as a time flow  $\Phi_T : \mathbb{R}^{N \times (1+P)} \times \mathbb{R} \rightarrow \mathbb{R}^{N \times (1+P)}$ . Therefore, the steady state  $\mathbf{X}_c^s$  is the fixed-point of the equation

$$\mathbf{X}_c = \Phi_T(\mathbf{X}_c). \quad (20)$$

When the above equation is not explicitly available, we can use the coarse time-stepper to approximate the time flow  $\Phi_T$ . Newton’s method or other iterative algorithms, often in matrix-free implementations, can be readily employed to compute steady-state gPC coefficients, which can be used to reconstruct random stable/unstable steady states of the original system, Eq. 14. It is also possible to compute limit cycles of the gPC coefficients through Poincaré return maps, thus linking EF methods with the analysis on random limit cycles.

**ACKNOWLEDGEMENTS** This work was partially supported by DARPA, and by an NSF Graduate Research Fellowship to K.B.



## References

1. Winfree, A. T. 2001. The geometry of biological time. volume 12 of *Interdisciplinary Applied Mathematics*. Springer-Verlag, New York. second edition.
2. Goldbeter, A. 1996. *Biochemical Oscillations and Cellular Rhythms: The Molecular Bases of Periodic and Chaotic Behaviour*. Cambridge University Press, Cambridge, UK.
3. Strogatz, S. H. 2003. *Sync: The Emerging Science of Spontaneous Order*. Hyperion, New York, NY.
4. Acebrón, J. A., L. L. Bonilla, C. J. Pérez Vicente, F. Ritort, and R. Spigler. 2005. The kuramoto model: A simple paradigm for synchronization phenomena. *Rev. Mod. Phys.* 77:137–185.
5. Strogatz, S. H. 2000. From Kuramoto to Crawford: exploring the onset of synchronization in populations of coupled oscillators. *Physica D* 143:1–20. Bifurcations, patterns and symmetry.
6. Forger, D. B., and C. S. Peskin. 2003. A detailed predictive model of the mammalian circadian clock. *PNAS* 100:14806–14811.
7. Michel, S., and C. S. Colwell. 2001. Cellular communication and coupling within the suprachiasmatic nucleus. *Chronobiol. Int.* 18:579–600.
8. Moon, S. J., and I. G. Kevrekidis. 2006. An equation-free approach to coupled oscillator dynamics: the kuramoto model example. *International Journal of Bifurcations and Chaos* 16:2043–2052.
9. Moon, S. J., R. Ghanem, and I. G. Kevrekidis. 2006. Coarse-graining the dynamics of coupled oscillators. *Phys. Rev. Lett.* 96.
10. Fishman, G. 1996. *Monte Carlo: concepts, algorithms, and applications*. Springer-Verlag, New York.
11. Schuler, M. L., and M. Domach. 1983. Mathematical models of the growth of individual cells. In *Foundations of Biochemical Engineering, American Chemical Society* 93–133.
12. Ghanem, R. G., and P. D. Spanos. 1991. *Stochastic Finite Elements: A Spectral Approach*. Springer-Verlag, New York.
13. Ghanem, R. 1998. Probabilistic characterization of transport in heterogeneous porous media. *Computer Methods in Applied Mechanics and Engineering* 158:3–4.
14. Maitre, O. L., O. Knio, M. Reagan, H. Najm, and R. Ghanem. 2001. A stochastic projection method for fluid flow. i: Basic formulation. *J. Comp. Phys.* 173:481–511.

15. Reagan, M., H. Najm, B. Debusschere, O. L. Maitre, O. Knio, and R. Ghanem. 2004. Spectral stochastic uncertainty quantification in chemical systems. *Combustion Theory and Modelling* 8:607–632.
16. Xiu, D. B., and G. E. Karniadakis. 2002. The wiener-askey polynomial chaos for stochastic differential equations. *SIAM J. Sci. Comput.* 24:619–644.
17. Deb, M. K., I. M. Babuška, and J. T. Oden. 2001. Solution of stochastic partial differential equations using Galerkin finite element techniques. *Comput. Methods Appl. Mech. Engrg.* 190:6359–6372.
18. Maitre, O. L., O. Knio, H. Najm, and R. Ghanem. 2004. Uncertainty propagation using Wiener-Haar expansions. *J. Comp. Phys.* 197:28–57.
19. Reagan, M., H. Najm, O. Knio, R. Ghanem, and O. Lemaitre. 2003. Uncertainty propagation in reacting-flow simulations through spectral analysis. *Combustion and Flame* 132:545–555.
20. Theodoropoulos, C., Y.-H. Qian, and I. G. Kevrekidis. 2000. Coarse stability and bifurcation analysis using time-steppers: A reaction-diffusion example. *PNAS* 97:9840–9843.
21. Kevrekidis, I. G., C. W. Gear, J. M. Hyman, P. G. Kevrekidis, O. Runborg, and C. Theodoropoulos. 2003. Equation-free, coarse-grained multiscale computation: enabling microscopic simulators to perform system-level analysis. *Commun. Math. Sci.* 1:715–762.
22. Kevrekidis, I. G., C. W. Gear, and G. Hummer. 2004. Equation-free: the computer-assisted analysis of complex, multiscale systems. *AIChE J.* 50:1346–1354.
23. Xiu, D. B., R. Ghanem, and I. G. Kevrekidis. 2005. An equation-free approach to uncertain quantification in dynamical systems. *IEEE Computing in Science and Engineering Journal (CiSE)* 7:16–23.
24. Henson, M. A., D. Muller, and M. Reuss. 2002. Cell population modeling of yeast glycolytic oscillations. *Biochem. J.* 368:433–446.
25. Wolf, J., and R. Heinrich. 2000. Effect of cellular interaction on glycolytic oscillations in yeast: A theoretical investigation. *Biochem. J.* 345:321–334.
26. Aon, M. A., S. Cortassa, H. V. Westerhoff, and K. van Dam. 1992. Synchrony and mutual stimulation of yeast cells during fast glycolytic oscillations. *J. General Microbiology* 138:2219–2227.
27. Dano, S., P. G. Sorensen, and F. Hynne. 1999. Sustained oscillations in living cells. *Nature* 402:320–322.
28. Das, J., and H. G. Busse. 1991. Analysis of the dynamics of relaxation type oscillation in glycolysis of yeast extracts. *Biophys. J.* 60:363–379.

29. Betz, A., and B. Chance. 1965. Phase relationship of glycolytic intermediates in yeast cells with oscillatory metabolic control. *Archives of Biochemistry and Biophysics* 109:585–594.
30. Das, J., and H.-G. Busse. 1985. Long term oscillations in glycolysis. *J. Biochem.* 97:719–727.
31. Kreuzberg, K., and W. Martin. 1984. Oscillatory starch degradation and fermentation in the green algae *chlamydomonas reinhardtii*. *Biochem. Biophys. Acta.* 799:291–297.
32. Tornheim, K. 1988. Fructose 2,6-bisphosphate and glycolytic oscillations in skeletal muscle extracts. *J. Biol. Chem.* 263:2619–2624.
33. Chance, B., J. R. Williamson, D. Jamieson, and B. Schoener. 1965. Properties and kinetics of reduced pyridine nucleotide fluorescence of the isolated and *in vivo* rat heart. *Biochem. J.* 341:357–377.
34. Ibsen, K. H., and K. W. Schiller. 1967. Oscillations of nucleotides and glycolytic intermediates in aerobic suspensions of Ehrlich ascites tumor cells. *Biochem. Biophys. Acta.* 799:291–297.
35. Ghosh, A. K., B. Chance, and E. K. Pye. 1971. Metabolic coupling and synchronization of NADH oscillations in yeast cell populations. *Archives of Biochemistry and Biophysics* 145:319–331.
36. Richard, P., B. M. Bakker, B. Teusink, K. van Dam, and H. V. Westerhoff. 1996. Acetaldehyde mediates the synchronization of sustained glycolytic oscillations in populations of yeast cells. *Eur. J. Biochem.* 235:238–241.
37. Richard, P., J. A. Diderich, B. M. Bakker, B. Teusink, K. van Dam, and H. V. Westerhoff. 1994. Yeast cells with a specific cellular make-up and an environment that removes acetaldehyde are prone to sustained glycolytic oscillations. *FEBS Letters* 341:223–226.
38. Bier, M., B. M. Bakker, and H. V. Westerhoff. 2000. How yeast cells synchronize their glycolytic oscillations: A perturbation analytic treatment. *Biophys. J.* 78:1087–1093.
39. Goldbeter, A., and R. Lefever. 1972. Dissipative structures for an allosteric model. Application to glycolytic oscillations. *Biophys. J.* 12:1302–1315.
40. Selkov, E. E. 1975. Stabilization of energy charge, generation of oscillation and multiple steady states in energy metabolism as a result of purely stoichiometric regulation. *Eur. J. Biochem.* 59:151–157.
41. Wolf, J., and R. Heinrich. 1997. Dynamics of two-component biochemical systems in interacting cells: Synchronization and desynchronization of oscillations and multiple steady states. *Biosystems* 43:1–24.
42. Hynne, F., S. Dano, and P. Sorensen. 2001. Full-scale model of glycolysis in *saccharomyces cerevisiae*. *Biophys. Chem* 94:121–163.

43. Wolf, J., J. Passarge, O. J. G. Somsen, J. L. Snoep, R. Heinrich, and H. L. Westerhoff. 2000. Transduction of intracellular and intercellular dynamics in yeast glycolytic oscillations. *Biophys. J.* 78:1145–1153.
44. Abramowitz, M., and I. Stegun. 1970. Handbook of Mathematical Functions. Dover Publications, Inc., New York.
45. Aït-Sahalia, Y. 2002. Maximum likelihood estimation of discretely sampled diffusions: a closed-form approximation approach. *Econometrica* 70:223–262.
46. Gear, C. W., and I. G. Kevrekidis. 2002. Projective methods for stiff differential equations: Problems with gaps in their eigenvalue spectrum. *SIAM Journal of Scientific Computing* 24:1091–1106.
47. Gear, C. W., and I. G. Kevrekidis. 2003. Telescopic projective methods for parabolic differential equations. *J. Comput. Phys.* 187:95–109.
48. Lee, S. L., and C. W. Gear. 2005. Second-order accurate projective integrators for multiscale problems. *UCRL-JRNL-212640*.
49. Hirsch, M. W., S. Smale, and R. L. Devaney. 2004. Differential equations, dynamical systems, and an introduction to chaos. volume 60 of *Pure and Applied Mathematics (Amsterdam)*. Elsevier/Academic Press, Amsterdam. second edition.
50. Kelley, C. T. 1995. Iterative Methods for Linear and Nonlinear Equations. SIAM.
51. Henson, M. A. 2004. Modeling the synchronization of yeast respiratory oscillations. *Journal of Theoretical Biology* 231:443–458.
52. To, T.-L., M. A. Henson, E. D. Herzog, and F. J. Doyle III. accepted for publication. A computational model for intercellular synchronization in the mammalian circadian clock. *Biophys. J.*
53. Gear, C. W., T. J. Kaper, I. G. Kevrekidis, and A. Zagaris. 2005. Projecting to a slow manifold: singularly perturbed systems and legacy codes. *SIAM J. Appl. Dyn. Syst.* 4:711–732 (electronic).
54. Belkin, M., and P. Niyogi. 2003. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation* 15:1373–1396.
55. Nadler, B., S. Lafon, R. R. Coifman, and I. G. Kevrekidis. 2006. Diffusion maps, spectral clustering and reaction coordinates of dynamical systems. *Appl. Comput. Harmon. Anal.* 21:113–127.
56. Gear, C. W., I. G. Kevrekidis, and C. Theodoropoulos. 2002. Coarse integration/bifurcation analysis via microscopic simulators: Micro-Galerkin methods. *Computers and Chemical Engineering* 26:941–963.

| <b>Leading eigenvalues</b> |                           |
|----------------------------|---------------------------|
| coarse eigenvalues         | full eigenvalues          |
| 1.000004                   | 1.000219                  |
| 0.942622                   | 0.9426798                 |
| 0.496275                   | 0.4962654                 |
| 0.328651                   | 0.3273728                 |
| $0.100267 \pm 0.068410i$   | $0.100250 \pm 0.0683621i$ |
| 0.126292                   | 0.126075                  |

Table 1: Selected, leading eigenvalues from the computations of the coarse level and of the fine level are shown. The eigenvalue at 1 corresponds to a neutrally stable direction along the limit cycle.

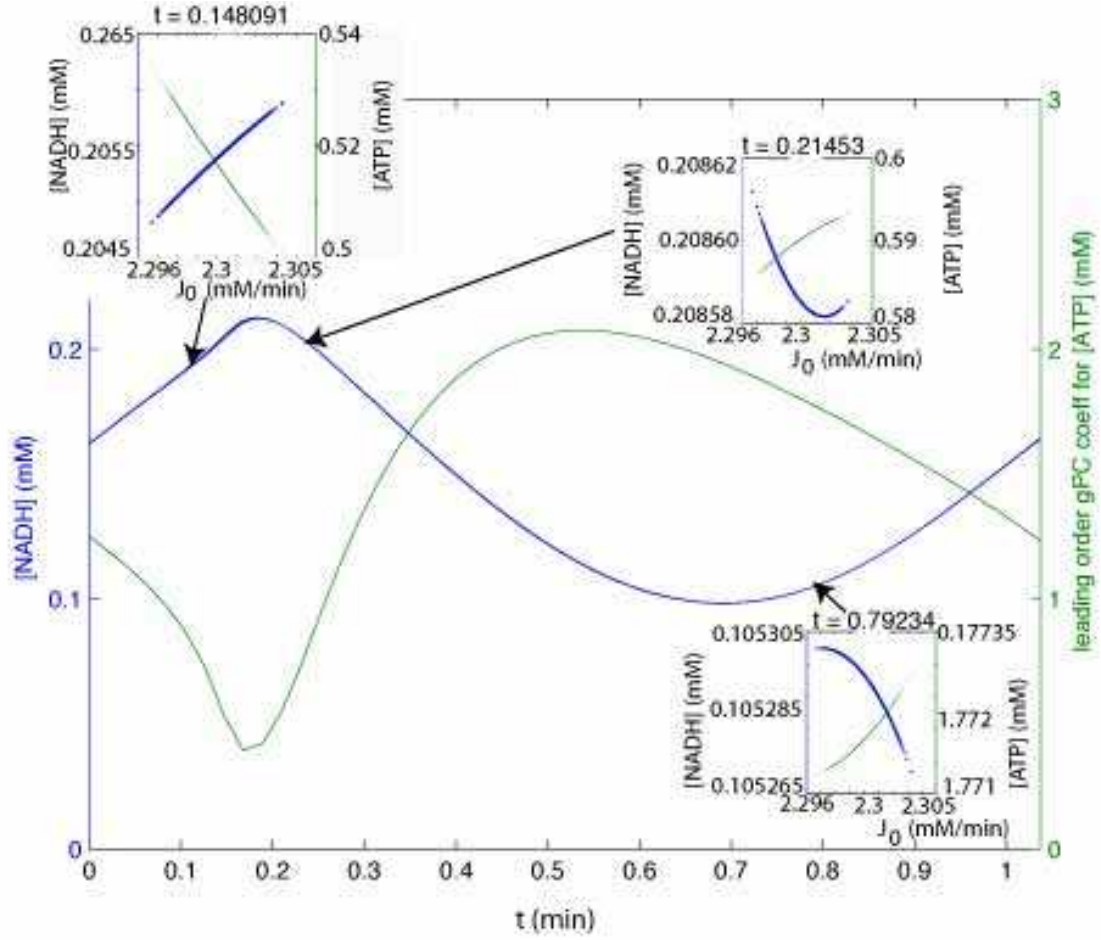


Figure 1: (color online) One period of the oscillation of [NADH] for an ensemble of yeast cells is shown (blue, left axis); one oscillation of the leading order gPC coefficient [ATP] is shown (green, right axis). The relationship of two intracellular concentrations, [NADH] (blue) and [ATP] (green), with respect to the heterogeneity of the glucose influx ( $J_0$ ) is shown in the inset figures. Note the continuous dependence of the intracellular concentration on the parameter  $J_0$ .



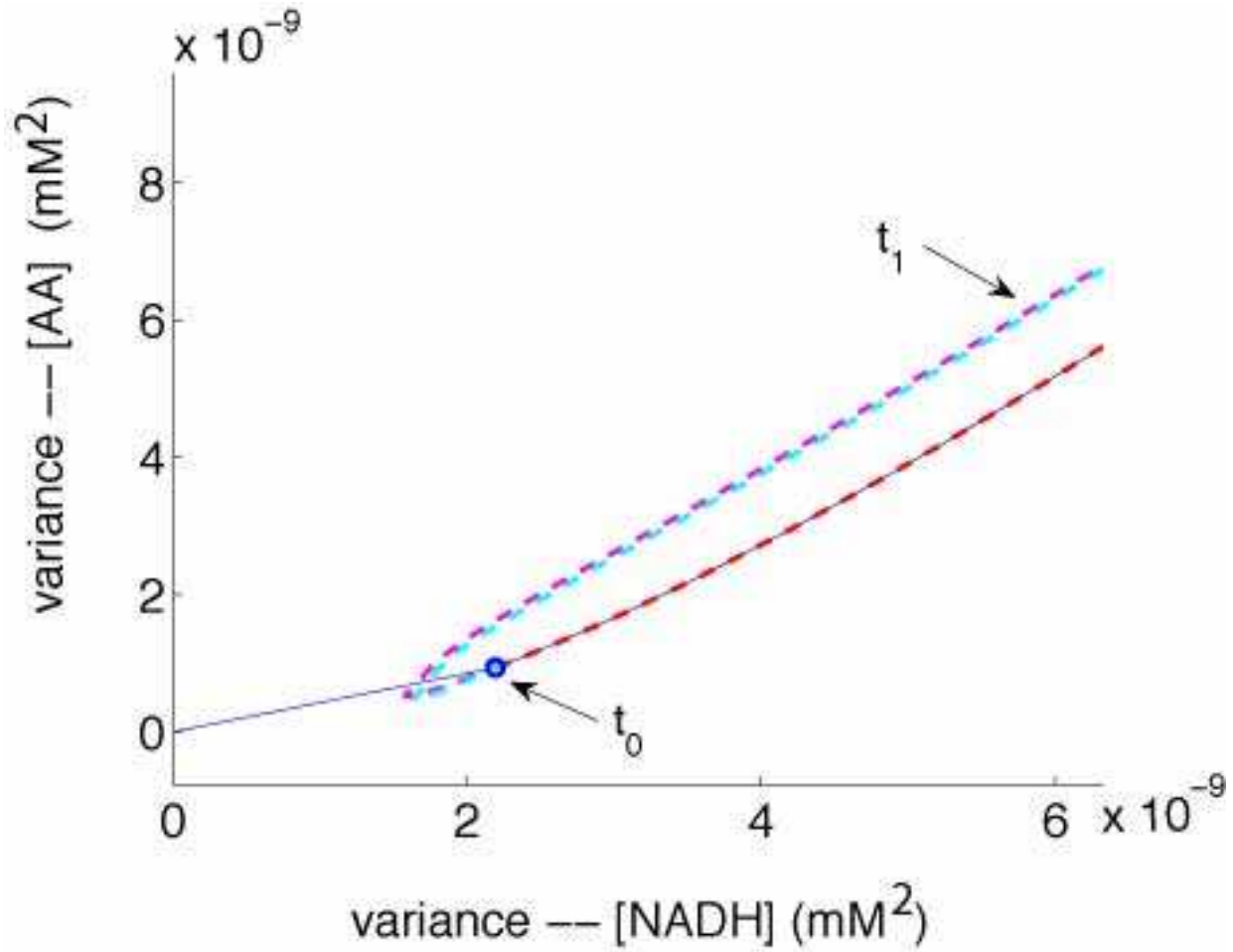


Figure 2: (color online) Phase maps for variances of [NADH] and [AA] in the initial stage. Blue curves: variances computed from realizations located on the fine-level limit cycle. Red curves: variances computed from realizations initialized with the fully correlated lifting from a “blue” coarse initial conditions. The blue circle indicates the location of this initial condition. Magenta and cyan dashed curves: variances computed from realizations initialized with two random liftings from the same coarse initial conditions as above. Note that blue and red curves almost coincide with each other.

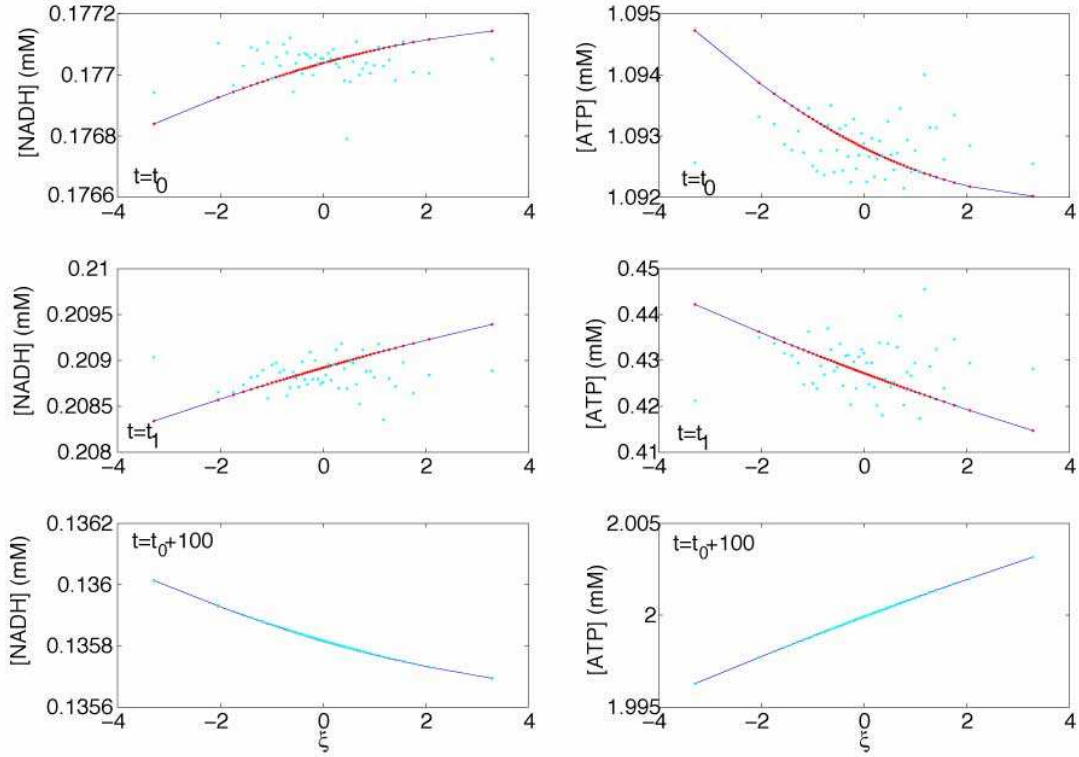


Figure 3: (color online) If the correlations of metabolites within a cell are not incorporated in the lifting (top: cyan dots at time  $t_0$ ), then these *wrong* correlations remain (middle,  $t_1$ ) until a sufficiently long amount of time elapses (bottom,  $t_1 + 100$ ). The blue line shows the mature, or natural, correlation. The red dots have been correctly initialized; the cyan points have been initialized with wrong initial correlations. For consistent computations, the relationships between the different metabolites in a cell must be replicated when an ensemble of cells is constructed (in the *lifting* step).

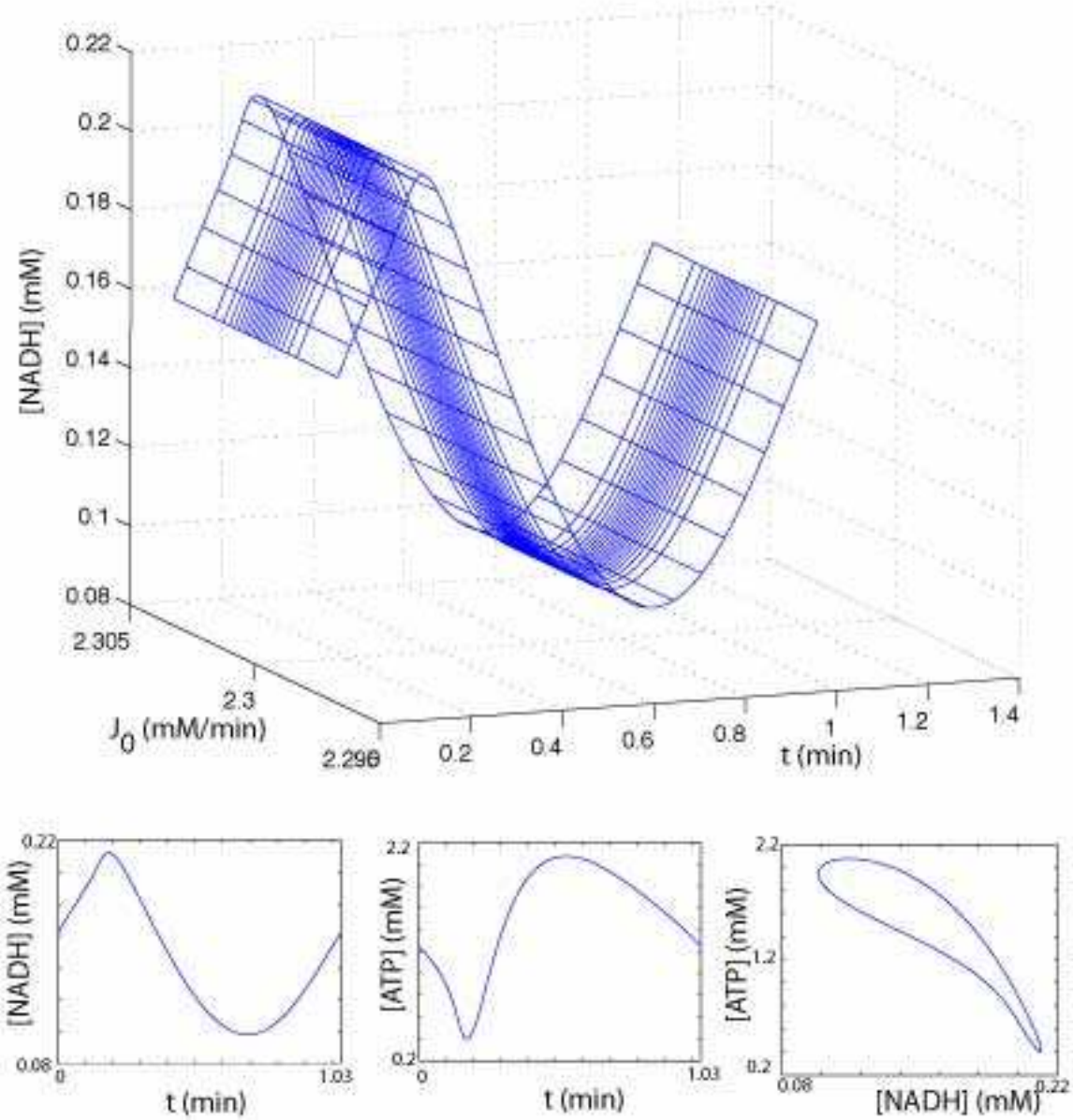


Figure 4: Top: The time series of [NADH] in cells of the ensemble are plotted against time  $t$  and the random parameter  $J_0$  (glucose influx) over one oscillation. The lines transverse to the time series curve are equal-time curves. Bottom left: Time series of the leading order gPC coefficients of [NADH] Bottom right: A projection of the coarse limit cycle onto the leading order [NADH] and [ATP] gPC coefficient plane. The distribution of the glucose influx parameter  $J_0$  has mean 2.3 and standard deviation 0.001.

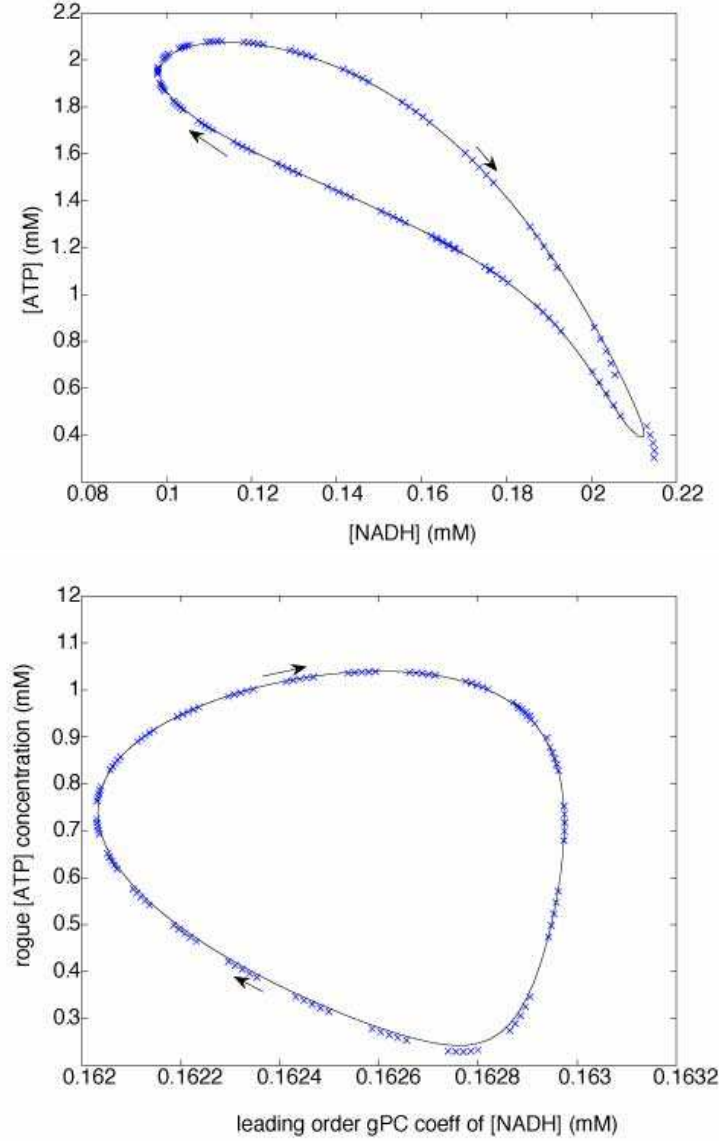


Figure 5: Projective integration in coarse variables (blue) and natural evolution (black). Projective integration accelerates computation of the evolution of a cell population. Top: An ensemble of size  $M = 1000$  with the distribution of the glucose influx parameter having mean 2.3 and standard deviation 0.001. Bottom: An ensemble of 50 cells, with the distribution of the glucose influx parameters having mean 2.1 and standard deviation 0.08. In this regime, there is one free cell, which oscillates at an amplitude relatively larger than the others. For each cell population subplot, the horizontal axis is the leading order gPC coefficient for [NADH]. Top: the vertical axis is the leading order gPC coefficient for [ATP]. Bottom: the vertical axis is [ATP] of the “rogue” cell.

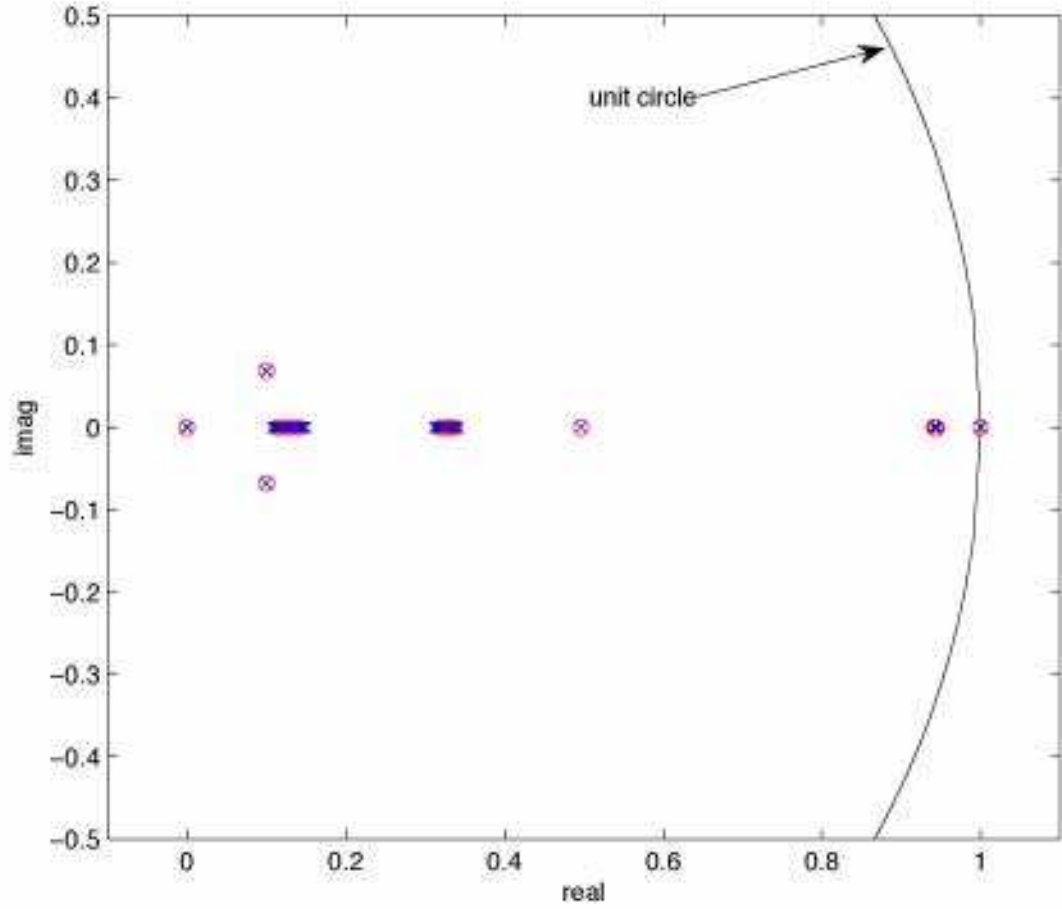


Figure 6: (color online) Leading eigenvalues of  $D\phi_t^f$  (blue,  $\times$ ) and  $D\phi_t$  (red,  $\circ$ ). The leading eigenvalues are essentially the same, as expected. the unit circle is shown (black curve); note the eigenvalues at 1, and all other eigenvalues are inside the unit circle. analyzing stability of the coarse problem gives insight into the stability of the full cell ensemble.

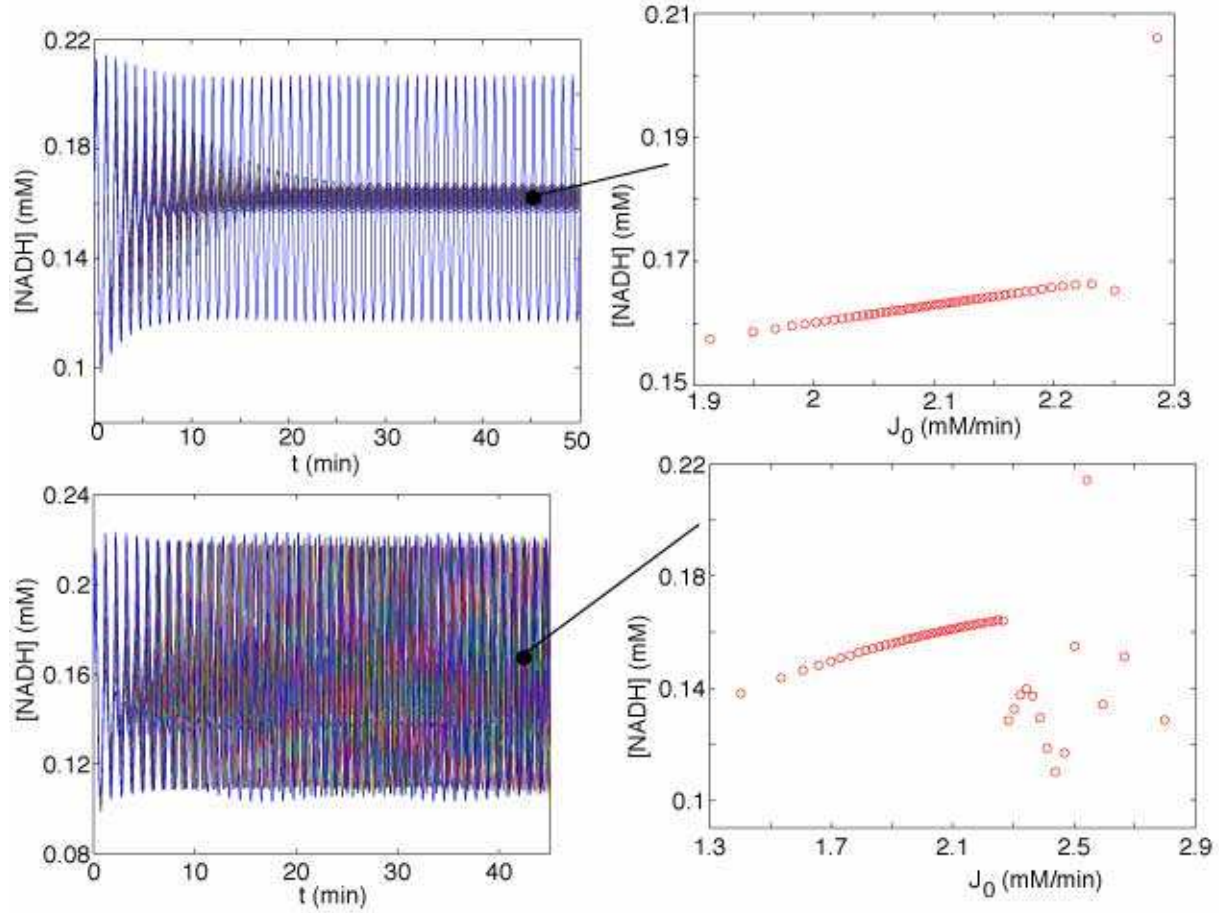


Figure 7: (color online) The discontinuity in the relationship between [NADH] and the parameter  $J_0$  is shown for distributions with “rogue” oscillators. Top:  $J_0$  has a Gaussian distribution with mean and standard deviation being 2.1 and 0.08, respectively; bottom: mean 2.1 and standard deviation 0.3. The figures on the left show the time history of [NADH] for all 50 cells. The figures on the right show [NADH] as a function of  $J_0$  at a snapshot in time ( $t = 45$ ). Equation-free computations can be applied in these regimes by using more and different coarse variables: gPC coefficients for the “bulk” of the oscillating cells and either gPC coefficients for the other cells or using all the state variables of the other cells.